จากตารางที่ 1 ได้รวบรวมผลงานวิจัยที่ผ่านมาในประเทศไทย จะเห็นได้ว่ามีทั้งการวิจัยระบบรู้จำเสียงแบบเฉพาะบุคคล และแบบไม่ขึ้นกับบุคคล แบบรู้จำเสียงคำเดี่ยว และในตารางทีที่ 2 เป็นงานวิจัย แบบรู้จำเสียงคำต่อเนื่อง ซึ่งมีทั้งที่เป็นคำพูด คำสั่งงานและตัวเลข แต่จะขออธิบายโดยละเอียดเฉพาะงานวิจัยที่มีลักษณะใกล้เคียงกัน ได้แก่

งานของ Areepongsa และ Jittapankul เป็นงานวิจัยแรกด้านการรู้จำเสียงพูดภาษาไทยที่นำเทคนิคแบบจำลองฮิดเดนมาคอฟโดยมาใช้โดยใช้แบบจำลองฮิดเดนมาร์คอฟแบบไม่ต่อเนื่อง จำนวน 3 สถานะ ทดลองกับตัวเลขคำเดี่ยวภาษาไทย ผลการรู้จำที่ได้คือ 82 %

ในงานของ Ahkuputra et al. (1997) ได้พัฒนาระบบรู้จำเสียงหลายพยางค์แบบไม่ขึ้นกับบุคคล มีคำศัพท์ในระบบทั้งหมด 70 คำ แบ่งเป็นคำศัพท์มีความยาว 1-3 พยางค์ อย่างละ 20 คำ และเป็นตัวเลข 0-9 10 คำ ใช้สัมประสิทธิ์การประมาณพันธะเชิงเส้น(LPC) 10 ลำดับ เป็นคุณลักษณะสำคัญ แบ่งพยางค์ด้วยวิธีการสร้างกราฟพลังงาน กำหนดค่าพลังงานเพื่อหาจุดเริ่มต้นและจุดสิ้นสุดของพยางค์ที่ 5% และ 10% จากพลังงานสูงสุดของเสียงต่อเนื่อง และสร้างแบบจำลองในระดับพยางค์ด้วยแบบจำลองฮิดเด็นมาร์คอฟแบบไม่ต่อเนื่อง แต่ละแบบจำลองมีทั้งหมด 15 สถานะ ได้อัตราการรู้จำเท่ากับ 86.75% คำศัพท์ยาว 1 พยางค์ 92.37% คำศัพท์ยาว 2 พยางค์ 86.25% สำหรับคำศัพท์ความยาว 3 พยางค์ และ 84.25% สำหรับตัวเลข ในงานวิจัยไม่ได้ระบุรายละเอียดเกี่ยวกับคำศัพท์ไว้ จึงไม่สามารถวิเคราะห์ระดับความคลุมเครือของระบบได้ ข้อดีของระบบนี้คือ สามารถรู้จำคำศัพท์ได้ในอัตราการรู้จำสูง เนื่องจากใช้แบบจำลองที่มีจำนวนสถานะมาก ทำให้ใช้เวลาในการฝึกฝนและรู้จำค่อนข้างนาน ทำให้ไม่สามารถนำมาใช้งานจริงได้ และทำให้การเพิ่มคำศัพท์ที่รู้จำในระบบเป็นไปได้ลำบาก

Nuttakorn Thubthong, Apirath Pusittrakul, Tanongkiat Sookawat and Boonserm Kijsirikul ในปี 2002 เสนอวิธีแก้ปัญหาการรู้จำเสียงวรรณยุกต์ที่ผิดเพี้ยน ในลักษณะเสียงกลืนกันของเสียง และเสียงมีความลาดเอียงด้วยแบบจำลองครึ่งเสียง(Half-Tone) เพื่อดึงข้อมูลจากพยางค์ข้างเคียงมาช่วยในการรู้จำ มีวิธีการคือ แบ่งเสียงต่อเนื่องออกเป็นพยางค์ด้วยมนุษย์โดยอาศัยการฟังและพิจารณารูปคลื่นเสียงประกอบกัน นำพยางค์ที่ได้มารู้จำโดยแบ่งออกเป็น 2 ส่วนเท่ากัน แต่ละส่วนคำนวณความถี่พื้นฐาน(F_0) และการเปลี่ยนแปลงตามเวลาของความถี่พื้นฐานเพื่อใช้เป็นค่าคุณลักษณะสำคัญดังนี้ ส่วนครึ่งแรกคำนวณความถี่พื้นฐานที่ตำแหน่ง 0, 10, 20, 30, 40 และ 50% และที่ตำแหน่ง 50, 60, 70, 80 และ 90% ของพยางค์ก่อนหน้า และส่วนครึ่งหลัง

คำนวณความถี่พื้นฐานที่ตำแหน่ง 50, 60, 70, 80, 90 และ 100% และที่ตำแหน่ง 10, 20, 30, 40 และ 50% ของพยางค์ถัดไป นำไปรู้จำด้วยโครงข่ายประสาทเทียมแบบเพอเซ็บตรอนหลายชั้น(multi-layer perceptron) ทดสอบด้วยเสียงผู้พูดทั้งหมด 10 คน แบ่งเป็นผู้ชาย 5 คน และผู้หญิง 5 คน พูดประโยคที่มีความยาว 4 พยางค์ทั้งหมด 11 ประโยค ผลการทดลองได้อัตราการรู้จำเท่ากับ 87.95% วิธีการดังกล่าวสามารถรู้จำเสียงวรรณยุกต์ต่อเนื่องได้ดี สามารถแก้ปัญหาการกลืนกันของเสียงวรรณยุกต์ได้ แต่ข้อจำกัดคือ ความถูกต้องในการรู้จำขึ้นกับความถูกต้องในการแบ่งพยางค์ เนื่องจากใช้การอ้างอิงขอบเขตพยางค์เพื่อหาค่าคุณลักษณะสำคัญ ดังนั้นวิธีการนี้จึงให้ผลดีเฉพาะระบบที่มีการแบ่งเสียงต่อเนื่องออกเป็นพยางค์ได้อย่างถูกต้องแล้ว

พิชัย ชูกาญจนพิทักษ์ ไชยันต์ สุวรรณชีวะศิริ และมนูญ พ่วงพูล ในปี 2002 เสนอการรู้จำเสียงพูดภาษาไทยต่อเนื่องแบบเฉพาะบุคคล มุ่งเน้นให้ระบบสามารถรู้จำคำศัพท์จำนวนมาก โดยใช้ลักษณะบ่งความต่างของหน่วยเสียง มีวิธีการคือ นำเสียงต่อเนื่องมาแบ่งส่วนเป็นคำศัพท์ด้วยการดูกราฟพลังงานสัมบูรณ์จากการวางกรอบสัญญาณขนาด 320 มิลลิวินาที หลังจากนั้นนำคำศัพท์มาหาตำแหน่งเริ่มต้นและสิ้นสุดของแต่ละพยางค์ ด้วยวิธีเดียวกันกับการแบ่งส่วนคำศัพท์ แต่เปลี่ยนขนาดของกรอบสัญญาณเป็น 9 มิลลิวินาที และมีส่วนสมมติจำนวนพยางค์ของคำศัพท์เพื่อช่วยให้ระบบมีการแบ่งพยางค์ที่ถูกต้องมากขึ้น ในกรณีที่เสียงพูดเกิดการซ้อนทับกันทำให้การใช้ระดับอ้างอิงไม่สามารถหารอยต่อระหว่างพยางค์ที่เหลื่อมกันได้ หลังจากนั้นนำพยางค์ที่แบ่งได้มาคำนวณคุณลักษณะสำคัญคือ สัมประสิทธิ์เซปสตรัมบนสเกลเมล(MFCC) พลังงาน เวลา และคาบเวลาพิตซ์ที่จุดต่าง ๆ โดยอ้างอิงจากขอบเขตพยางค์ ฝึกฝนโครงข่ายประสาทเทียมด้วยเสียงพูดต้นแบบคำศัพท์ 5 ชุด ๆ ละ 1,000 คำ และทดสอบเสียงพูดคำศัพท์ในกลุ่มทดสอบที่มีจำนวนเท่ากันกับกลุ่มต้นแบบ ผลการทดลอง การรู้จำหน่วยเสียงพยัญชนะต้น สระ พยัญชนะสะกด และวรรณยุกต์ เท่ากับ 56.24%, 79.84%, 60.32% และ 82.16% นำไปรู้จำคำศัพท์ในฐานข้อมูลคำศัพท์ได้ความถูกต้องเฉลี่ยเท่ากับ 64.22% เมื่อใช้จำนวนแบบหน่วยเสียง 1 แบบ/คำศัพท์ และ 84.28% เมื่อใช้จำนวนแบบหน่วยเสียง 6 แบบ/คำศัพท์ วิธีการดังกล่าวมีข้อดีคือ สามารถเพิ่มจำนวนคำศัพท์ให้มากขึ้นได้ โดยเพิ่มในฐานข้อมูลคำศัพท์และแบบหน่วยเสียง โดยไม่ต้องสอนโครงข่ายประสาทเทียมใหม่ แต่ยังมีข้อจำกัดคือ จากผลการทดลองจะเห็นว่าความถูกต้องของการรู้จำยังให้ผลที่ยังไม่ค่อยดีนัก และการหาคุณลักษณะสำคัญของหน่วยเสียงจำเป็นต้องทราบตำแหน่งของหน่วยเสียง ซึ่งเป็นผลจากการแบ่งส่วนพยางค์และการแบ่งส่วนคำศัพท์ ดังนั้นตำแหน่งที่นำไปหาคุณลักษณะสำคัญของหน่วยเสียงจึงมีผลต่อการรู้จำหน่วยเสียงโดยตรง

ฐานียา, 2546 ได้ทำการรู้จำเสียงพูดตัวเลขแบบต่อเนื่องเพื่อใช้ในการรู้จำเสียงพูดเฉพาะบุคคลสำหรับการใช้งานอีเมล์ โดยใช้วิธีการรู้จำที่ไม่มีการแบ่งเสียงต่อเนื่องเป็นพยางค์ อีเมล์เป็นแอพพลิเคชันที่มีคำสั่งงานแบ่งได้เป็น 2 ลักษณะคือ คำสั่งงานแบบจำกัด และคำสั่งงานแบบไม่จำกัด คำสั่งงานแบบจำกัด เป็นคำสั่งงานหลักของแอพพลิเคชัน เช่น ลบจดหมาย อ่านจดหมาย เป็นต้น คำสั่งงานแบบไม่จำกัด เป็นคำสั่งงานซึ่งขึ้นกับผู้ใช้งาน เช่น อีเมล์แอดเดรส ชื่อไฟล์ เป็นต้น โดยในงานวิจัยนี้ได้กำหนดคำสั่งงานหลักภาษาไทยต่อเนื่องทั้งหมด 23 คำสั่ง เมื่อทดสอบกับผู้ใช้ทั้งหมด 12 คนได้อัตราการรู้จำเสียงเฉลี่ยเท่ากับ 87.88%

อุปกรณ์และวิธีการ

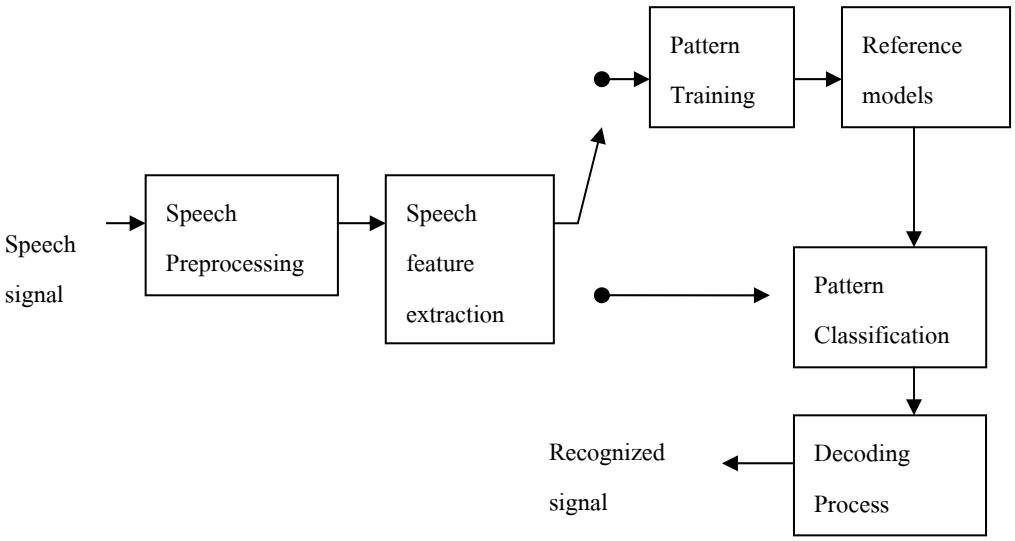
อุปกรณ์

1. เครื่องคอมพิวเตอร์ Intel Celeron 1.3 กิกกะเฮิรตซ์ หน่วยความจำ 128เมกกะไบต์
2. ไมโครโฟน
3. ลำโพง หรือหูฟัง
4. ตัวแปลภาษา g++
5. ซอฟต์แวร์ช่วยพัฒนา K Develop
6. ระบบปฏิบัติการ Linux

วิธีการ

ระบบรู้จำเสียงพูดในงานวิจัยนี้ ประกอบด้วย 5 ส่วนหลัก คือ การประมวลผลสัญญาณเสียงพูดเบื้องต้น การดึงค่าคุณลักษณะสำคัญเสียงพูด การฝึกฝนระบบ และ ส่วนการรู้จำ ซึ่งได้อธิบายทฤษฎีไว้ในหัวข้อความรู้พื้นฐาน ในบทการตรวจเอกสาร งานวิจัยนี้ได้เลือกเทคนิควิธีการในแต่ละส่วนที่เหมาะสม เพื่อสร้างระบบรู้จำเสียงต่อเนื่องแบบไม่ขึ้นกับบุคคล

กระบวนการรู้จำเสียงของระบบรู้จำเสียงต่อเนื่องตัวเลขภาษาไทย แบ่งเป็น 2 ขั้นตอนคือ ขั้นตอนการฝึกฝน และขั้นตอนการรู้จำ ดังภาพที่ 14 มีรายละเอียดดังต่อไปนี้



ภาพที่ 14 กระบวนการรู้จำเสียง

1. การกำหนดเสียงตัวอย่าง

ข้อมูลเสียงที่ใช้ในงานวิจัยนี้เป็นเสียงของผู้พูดอายุระหว่าง 20-31 ปี โดยผู้พูดแต่ละคนพูดเสียงตัวเลขภาษาไทยความยาว 7 คำต่อเสียงพูด จำนวนคนละ 20 ชุดของตัวเลขเพื่อให้ครอบคลุมความเป็นไปได้ในการเชื่อมต่อของตัวเลขแต่ละตัวรายละเอียดของคำเรียกในการบันเสียงแสดงไว้ในภาคผนวก ก

หลักเกณฑ์ในการบันทึกเสียง

1. บันทึกเสียงพูดทางโทรศัพท์เข้าทางการ์ดเสียง โดยมีอัตราการชักตัวอย่าง (Sampling Rate) ที่ 8,000 เฮิรตซ์ ขนาด 8 บิต
2. บันทึกเสียงภายใต้สภาพแวดล้อมของบ้านและที่ทำงาน
3. ผู้พูดเป็นผู้ที่มีการออกเสียงปกติและใช้สำเนียงกรุงเทพเป็นภาษาพูด

ข้อมูลเสียงถูกแบ่งเป็นสองชุดคือ ชุดฝึกฝน (Training Set) และชุดทดสอบ (Testing Set) ชุดฝึกฝนประกอบด้วยเสียงของผู้พูดจำนวน 100 คนแบ่งเป็นชายจำนวน 50 คน หญิงจำนวน 50 คน อายุระหว่าง 20-31 ปี จำนวนของชุดเสียงพูดทั้งหมดนำไปสร้างและฝึกฝนระบบคือ 2,000

เสียงพูด ในการส่วนชุดทดสอบซึ่งเป็นคนละกลุ่มผู้พูด จำนวน 40 คนเป็นชาย 20 คนและหญิง 20 คน

2. ขั้นตอนการฝึกฝน Training Procedure

ขั้นตอนการฝึกฝนประกอบด้วยขั้นตอนย่อย 4 ขั้นตอน แต่ละขั้นตอนมีรายละเอียดดังต่อไปนี้

2.1 การประมวลผลสัญญาณเสียงพูดเบื้องต้น

การประมวลผลสัญญาณเสียงพูดเบื้องต้นแบ่งเป็น 2 ขั้นตอนคือ

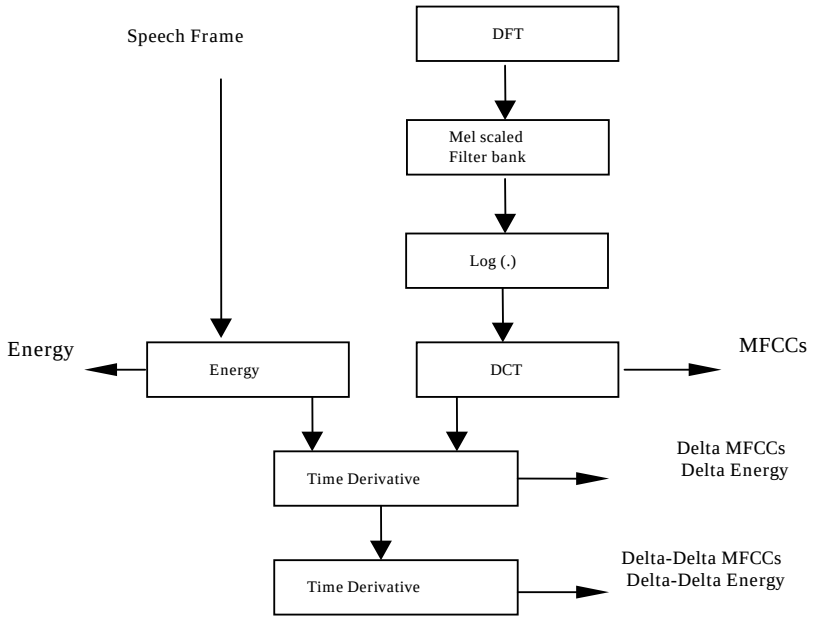
2.1.1 ขั้นตอนกรรมวิธีการเน้นล่วงหน้า นำเสียงที่ผ่านการบันทึกมาผ่านกรรมวิธีการเน้นล่วงหน้าเพื่อปรับสเปกตรัมให้มีความเรียบและเพิ่มอัตราส่วนของสัญญาณเสียงพูดต่อสัญญาณรบกวน

2.1.2 ขั้นตอนการวางกรอบสัญญาณ โดยเลือกใช้กรอบสัญญาณแบบแฮมมิง ดังสมการ (1) มีลักษณะดังภาพที่ (1) โดยกำหนดความกว้างของกรอบเท่ากับ 30 มิลลิวินาที และเลื่อนกรอบสัญญาณครั้งละ 10 มิลลิวินาที การวางกรอบสัญญาณเพื่อทำให้การเปลี่ยนแปลงสัญญาณเป็นไปอย่างช้าๆหลีกเลี่ยงความไม่ต่อเนื่องของสัญญาณ

2.2 การดึงค่าคุณลักษณะสำคัญของเสียงพูด

ค่าคุณลักษณะสำคัญที่เลือกใช้ คือ สัมประสิทธิ์เซปสตรัมบนสเกลเมล(MFCC) คำนวณได้ดังสมการ (11) และ การเปลี่ยนแปลงสัมประสิทธิ์เซปสตรัมบนสเกลเมล(Differential MFCC หรือ DMFCC) อันดับที่ 1 และ อันดับที่ 2 คำนวณได้ดังสมการที่ (12)

ในการเลือกใช้ สัมประสิทธิ์เซปสตรัมบนสเกลเมล(MFCC) เนื่องมาจาก มีคุณสมบัติในการขยายความชัดเจนของเสียงในช่วงความถี่สูงซึ่งมีลักษณะเหมือนกับการรับฟังเสียงของมนุษย์ โดยจำนวนสัมประสิทธิ์ที่เหมาะสมนั้น

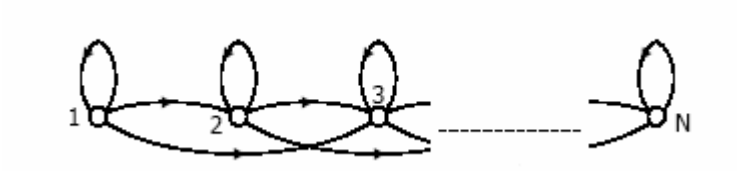


ภาพที่ 13 การดึงคาคุณลักษณะสำคัญของเสียง

ภาพที่ 13 เป็นรายละเอียดการหาค่าตุณลักษณะสำคัญของเสียง โดยอินพุจะเป็นเสียงพูดที่ผ่านการวางกรอบสัญญาณแล้วหลังจากนั้นทำการหาค่าในโดเมนความถี่โดยใช้การเปลี่ยนแปลงสัญญาณด้วยดิสครีทฟูเรียร์ แล้วทำการหาส่วนความถี่ที่หูมนุษย์รับฟังได้ด้วยฟังก์ชันความถี่สเกลเมล หลังจากนั้นใช้การแปลงสัญญาณดิสครีทโคซายน์จะได้ผลลัพธ์สัมประสิทธิ์เซปสตรัมบนสเกลเมลออกมา

2.3 การฝึกฝน

ใช้ทฤษฎีแบบจำลองฮิดเดนมาร์คอฟ เพื่อฝึกฝนและรู้จำ โดยได้ใช้แบบจำลองแบบซ้ายไปขวาทำการเปลี่ยนจำนวนสถานะเพื่อหาค่าที่เหมาะสม ซึ่งเป็นกับเสียงพูดที่กำหนดให้ผู้พูดฝึกฝนระบบ ดังนั้นจึงไม่ได้มีส่วนที่ทำหน้าที่แบ่งเสียงต่อเนื่องออกเป็นพยางค์ แบบจำลองในระดับพยางค์มีลักษณะดังต่อไปนี้



ภาพที่ 14 แบบจำลองของแต่ละเสียงพูด

การฝึกฝนเสียงพูดในระดับพยางค์ใด ๆ นั้น ตามอัลกอริทึมฟอร์เวิร์ดแบคเวิร์ด (Forward-Backward) โดยใช้สัญลักษณ์ที่กำหนดขึ้น12 ค่า เป็นตัวเลข 0 ถึง 9 สัญลักษณ์อีกสองค่า คือ ช่วงเสียงเงียบ(Silence) และ ช่วงหยุดระหว่างการพูด (Pause)

หลังจากฝึกฝนแบบจำลองพยางค์หนึ่ง ๆ จะได้ค่าความน่าจะเป็นประจำแบบจำลอง(a, b และ π) เก็บไว้เพื่อใช้ในการรู้จำ

3. ขั้นตอนการรู้จำ

หลังจากผ่านขั้นตอนการฝึกฝนแล้ว ภายในระบบจะมีชุดแบบจำลอง และฐานข้อมูลที่เก็บไว้เพื่อให้ระบบสามารถรู้จำได้ไว้ ขั้นตอนการรู้จำประกอบด้วยขั้นตอนย่อย 4 ขั้นตอน ประกอบด้วย

1) การประมวลสัญญาณเสียงพูดเบื้องต้น
2) การดึงค่าคุณลักษณะสำคัญ
3) การรจำแนกรูปแบบหรือการรู้จำคำต่อเนื่อง
4) การหาคำตอบ

ขั้นตอนย่อย 2 ขั้นตอนแรกคือ คือ การประมวลผลสัญญาณเสียงพูดเบื้องต้น การดึงค่าคุณลักษณะสำคัญเสียงพูด นั้น มีวิธีการเดียวกับในขั้นตอนการฝึกฝน ดังนั้นจึงขออธิบายรายละเอียดขั้นตอนการรู้จำเฉพาะ 2 ขั้นตอนย่อยที่เหลือคือ การรู้จำคำต่อเนื่อง และการหาคำตอบคำศัพท์ ซึ่งมีรายละเอียดดังต่อไปนี้

3.1 <u>การรู้จำคำต่อเนื่อง</u>

ในการรู้จำทำได้โดยนำขั้นตอนวิธีแบบฟอร์เวิร์ด ดังหัวข้อ 3.1.2 ในบทการตรวจเอกสาร เพื่อหาค่าความจะเป็นสูงสุดที่จะให้ผลลัพธ์เป็นตัวเลขแต่ละตัว

อัลกอริธึมฟอร์เวิร์ด เป็นอัลกอริธึมที่ใช้ในการรู้จำ โดยการหาค่าความน่าจะเป็นที่แบบจำลอง(λ) ให้ลำดับของผลลัพธ์($O=O_1, O_2, \ldots, O_T$) $P(O \mid \lambda)$ ดังสมการ (15)

$$P(O \mid \lambda) = \sum_{i=1}^{N} \alpha_T(i) \tag{15}$$

$$\alpha_{t+1}(i) = \begin{cases} \pi_i b_i(O_1), & 1 \le i \le N, t = 0 \\ \left[\sum_{j=1}^{N} \alpha_t(j) a_{ji} \right] b_i(O_{t+1}), & 1 \le j \le N, 1 \le t \le T-1 \end{cases} \tag{16}$$

3.2 <u>การหาคำตอบ</u>

การรู้จำคำศัพท์ มีหน้าที่ในการหาคำตอบที่ถูกต้องที่สุดจากชุดตัวเลขที่ได้จากการรู้จำตัวเลขต่อเนื่อง รู้จำคำศัพท์โดยใช้ค่าความน่าจะเป็นซึ่งนำข้อมูลสถิติในฐานข้อมูลคำศัพท์มาประกอบการเลือกคำตอบใช้ขั้นตอนวิธีไวเทอร์บิในการหาคำตอบ อัลกอริธึมไวเทอร์บิ เป็นอัลกอริธึมที่ใช้ในการหาลำดับของสถานะ($Q=\{q_1, q_2, \ldots, q_T\}$) ที่ให้ค่าความน่าจะเป็นสูงที่สุด(P^*) ของแบบจำลอง มีประโยชน์ในการรู้จำเสียงที่ประกอบด้วยแบบจำลองหลาย ๆ แบบจำลองต่อกัน ลำดับของสถานะที่ให้ค่าความน่าจะเป็นสูงที่สุดหาได้จากสมการ (17)

$$q_t^* = \psi_{t+1}(q_{t+1}^*), \qquad t = T-1, T-2, \ldots, 1 \tag{17}$$

และค่าความน่าจะเป็นสูงที่สุดหาได้จากสมการ (18)

$$P^* = \max_{1 \le i \le N} \left[\delta_T(i) \right] \tag{18}$$

โดย Ψ_t คือ สถานะที่เวลาใด ๆ หาได้จากสมการ (19)

$$\psi_t(j) = \begin{cases} 0, & 1 \le j \le N, t = 1 \\ \arg\max\limits_{1 \le i \le N} [\delta_{t-1}(i) a_{ij}], & 1 \le j \le N, 2 \le t \le T \end{cases} \quad (19)$$

และ δ_t คือ ค่าความน่าจะเป็นที่เวลาใด ๆ หาได้จากสมการ (20)

$$\delta_t(j) = \begin{cases} \pi_j b_j(O_1), & 1 \le j \le N, t = 1 \\ \max\limits_{1 \le i \le N} \left[\delta_{t-1}(i) a_{ij}\right] b_j(O_t), & 1 \le j \le N, 2 \le t \le T \end{cases} \quad (20)$$

จากขั้นตอนวิธีที่การรู้จำคำศัพท์จะได้ลำดับที่ดีที่สุดของความน่าจะเป็นของชุดตัวเลขแบบต่อเนื่องโดยใช้สมการที่ 20

ผลและวิจารณ์

ในบทนี้ประกอบ วิธีการวัดประสิทธิภาพของระบบ และ ผลการทดลองของระบบรู้จำเสียงพูดภาษาไทยต่อเนื่องตัวเลขภาษาไทย โดยในส่วนของผลการทดลองจะแสดงให้เห็นถึงผลการทดลองใน 5 หัวข้อ คือ 1) ผลการทดลองเพื่อหาจำนวนค่าคุณลักษณะที่เหมาะสม 2) ผลการทดลองหาจำนวนสถานะที่เหมาะสมและกลุ่มเกาส์เซียนของแบบจำลองฮิดเดนมาคอร์ฟ 3) ผลการทดลองจำนวนเสียงพูดที่ใช้ในการฝึกฝนรูปแบบ และ 4) ผลการวิเคราะห์ค่าความผิดพลาด

วิธีการวัดประสิทธิภาพของระบบ

ในการวัดประสิทธิภาพทำได้โดยการเทียบผลลัพธ์ที่ได้จากระบบกับคำตอบที่ต้องการ ผลที่ได้จากการเทียบคืออัตราความถูกต้องของการรู้จำคำ (Word accuracy หรือ WER) ดังสมการที่ 26

$$\text{อัตราความถูกต้องของการรู้จำคำ} = 1 - \text{อัตราข้อผิดพลาด} \quad (26)$$

โดยวิธีการวัดประสิทธิภาพนี้ใช้หลักการเดียวกับงานวิจัยด้านการรู้จำเสียงพูดในหลายๆงาน เช่น งานของ Lee, 1989; Bo *et al,* 1999 และ Hansen *et al*., 2001 เป็นต้น จากในสมการที่ 26 จะต้องทำการหาค่าอัตราข้อผิดพลาดให้ได้จึงจะสามารถคำนวณหาอัตราการรู้จำคำ การคำนวณหาอัตราข้อผิดพลาดดังสมการที่ 27

$$\text{อัตราข้อผิดพลาด} = \frac{\text{ผลรวมข้อผิดพลาดทั้งหมด}}{\text{จำนวนคำศัพท์ที่ทดสอบทั้งหมด}} \quad (27)$$

ข้อผิดพลาดที่เกิดขึ้นจากการทดลองสามารถแบ่งเป็น 3 ชนิด คือ

1. ข้อผิดพลาดแบบแทรก (Insertion errors) เกิดขึ้นเมื่อผลลัพธ์ที่ได้มาไม่มีความเกี่ยวข้องกับคำตอบอ้างอิง เช่น คำตอบอ้างอิงคือ 7654231 ผลลัพธ์ที่ได้คือ 71654321 ข้อผิดพลาดที่เกิดขึ้นนี้คือระหว่างตำแหน่งที่หนึ่งและสอง (เลข7 และ 6) ของคำตอบอ้างอิงถูกแทรกตัวเลขมาหนึ่งตัวเลข (เลข 1)

2. ข้อผิดพลาดแบบแทนที่ (Substitution errors) เกิดขึ้นเมื่อเทียบคำตอบอ้างอิงและผลลัพธ์แล้วพบว่าผลลัพธ์ที่ได้ไม่ถูกต้องในรูปแบบการถูกแทนที่เช่น คำตอบอ้างอิงคือเลข 7654321 แต่ผลลัพธ์ที่ค้มาคือเลข 7154321 ตัวเลขที่ผิดไปคือตัวเลขในตำแหน่งที่สอง คือเลข 6 ถูกแทนที่ด้วยเลข 1

3. ข้อผิดพลาดแบบลบหาย (Deletion errors) จะเกิดขึ้นเมื่อคำตอบอ้างอิงไม่มีความเกี่ยวข้องกับผลลัพธ์ในลักษณะที่มีตัวเลขในคำตอบอ้างอิงหายไป เช่น คำตอบอ้างอิงคือ 7654321 แต่ผลลัพธ์ที่ได้คือ 754321ข้อผิดพลาดแบบลบหายนี้คือตัวเลขในตำแหน่งที่สองหายไป (เลข 6)

ตารางที่ 4 ตัวอย่างผลลัพธ์ตัวเลข 9989796 และค่าข้อผิดพลาดที่เกิดขึ้น

ผลลัพธ์ของตัวเลขที่ได้	ข้อผิดพลาดแบบแทรก	ข้อผิดพลาดแบบแทนที่	ข้อผิดพลาดแบบลบหาย
9989796	0	0	0
965989796	2	0	0
9983796	0	1	0
959796	0	1	1
9987796	0	1	0
9589796	0	1	0
89689796	1	1	0
9985796	0	1	0
9929796	0	1	0
9968876896	3	1	0

ตารางที่ 4 เป็นตัวอย่างผลลัพธ์ตัวเลข 9989796 และค่าข้อผิดพลาดที่เกิดขึ้นโดยทำการทดสอบตัวเลขจำนวน 10 ชุดตัวเลข ตัวเลขแต่ละตัวมีความยาว 7 คำเรียงติดกันจำนวนคำศัพท์ที่ทดสอบทั้งหมดคือ 70 คำ ค่าข้อผิดพลาดที่เกิดขึ้นคือ 16 ข้อผิดพลาด จากสมการที่ 27 จะได้อัตราข้อผิดพลาดคือ 16/70 หรือ 22.85 % และจาอัตราค่าความถูกต้องเท่ากับ 77.15%

ผลการทดลอง

ในงานวิจัยนี้ได้ทดสอบประสิทธิภาพในการรู้จำ โดยแบ่งเป็นทั้งหมด 4 ผลการทดลอง ประกอบคือ 1) ผลการทดลองเพื่อหาจำนวนค่าคุณลักษณะที่เหมาะสม 2) ผลการทดลองหาจำนวนสถานะและกลุ่มเกาส์เซียนที่เหมาะสมของแบบจำลองฮิดเดนมาคอร์ฟ 3) ผลการทดลองจำนวนเสียงพูดที่ใช้ในการฝึกฝนรูปแบบ และ 4) ผลการวิเคราะห์ค่าความผิดพลาด สำหรับผลการทดลองที่ 1-3 นั้นเป็นส่วนของการทดลองขั้นตอนวิธีของระบบโดยผลการทดลองที่ 1 เป็นขั้นตอนวิธีการดึงค่าคุณลักษณะสำคัญของเสียงเพื่อหาค่าคุณลักษณะที่ดีที่สุดต่อระบบ ส่วนผลการทดลองที่ 2 และ 3 เป็นการทดลองหาค่าพารามิเตอร์ที่เหมาะสมเพื่อใช้ในขั้นตอนการจำแนกรูปแบบและการฝึกฝนรูปแบบ

1. <u>ผลการทดลองเพื่อหาจำนวนค่าคุณลักษณะที่เหมาะสม</u>

กรรมวิธีการดึงค่าคุณลักษณะสำคัญของเสียงเป็นการตัวแทนของเสียงพูดเพื่อการรู้จำซึ่งสามารถมีได้หลายคุณลักษณะ ค่าของตัวแทนเสียงที่ดีจะช่วยให้การจำแนกรูปแบบมีประสิทธิภาพที่ดีขึ้นด้วย ในงานวิจัยนี้เลือกใช้ 6 คุณลักษณะหลัก คือ สัมประสิทธิ์เซปสตรัมบนสเกลเมล (MFCC) ผลต่างทางเวลาของสัมประสิทธิ์เซปสตรัมบนสเกลเมลลำดับที่หนึ่ง (ΔMFCC) ผลต่างทางเวลาของสัมประสิทธิ์เซปสตรัมบนสเกลเมลลำดับที่สอง (ΔΔMFCC) ค่าพลังงานเสียง(E) ผลต่างค่าพลังงานเสียงในลำดับที่หนึ่ง (ΔE) และ ผลต่างค่าพลังงานเสียงในลำดับที่สอง (ΔΔE) ในตารางที่ 5 แสดงรายละเอียดค่าคุณลักษณะที่ใช้

สำหรับการทดสอบเพื่อหาค่าคุณลักษณะที่เหมาะสมนั้นทำการทดลอง 8 ชุดการทดลอง โดยรายละเอียดการทดลองในแต่ละชุดของการดึงค่าคุณลักษณะสำคัญของเสียงพูดมีดังนี้

1. การทดลองชุดที่หนึ่ง (F1) 26 สัมประสิทธิ์ ประกอบด้วยค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมล (MFCC) จำนวน 13 สัมประสิทธิ์ และผลต่างค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมลลำดับที่หนึ่ง (ΔMFCC) จำนวน13 สัมประสิทธิ์

2. การทดลองชุดที่สอง (F2) 27 สัมประสิทธิ์ ประกอบด้วยค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมล (MFCC) จำนวน 13 สัมประสิทธิ์ ผลต่างค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมลลำดับที่หนึ่ง (ΔMFCC) จำนวน13 สัมประสิทธิ์ และค่าผลต่างพลังงาน (ΔE) 1 สัมประสิทธิ์

3. การทดลองชุดที่สาม (F3) 28 สัมประสิทธิ์ ประกอบด้วยค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมล (MFCC) จำนวน 13 สัมประสิทธิ์ และผลต่างค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมลลำดับที่หนึ่ง (ΔMFCC) จำนวน13 สัมประสิทธิ์ ค่าผลต่างพลังงาน (ΔE)และค่าผลต่างพลังงานลำดับที่สอง (ΔΔE) อย่างละ 1 สัมประสิทธิ์

4. การทดลองชุดที่สี่ (F4) 29 สัมประสิทธิ์ ประกอบด้วยค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมล (MFCC) จำนวน 13 สัมประสิทธิ์ และผลต่างค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมลลำดับที่หนึ่ง (ΔMFCC) จำนวน13 สัมประสิทธิ์ ค่าพลังงาน (E) ค่าผลต่างพลังงาน (ΔE) และค่าผลต่างพลังงานลำดับที่สอง (ΔΔE)

5. การทดลองชุดที่ห้า (F5) 39 สัมประสิทธิ์ ประกอบด้วยค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมล (MFCC) จำนวน 13 สัมประสิทธิ์ ผลต่างค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมลลำดับที่หนึ่ง (ΔMFCC) จำนวน13 สัมประสิทธิ์ และ ผลต่างค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมลลำดับที่สอง (ΔΔMFCC) จำนวน13 สัมประสิทธิ์

6. การทดลองชุดที่หก (F6) 40 สัมประสิทธิ์ ประกอบด้วยค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมล (MFCC) จำนวน 13 สัมประสิทธิ์ ผลต่างค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมลลำดับที่หนึ่ง (ΔMFCC) จำนวน13 สัมประสิทธิ์ ผลต่างค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมลลำดับที่สอง (ΔΔMFCC) จำนวน13 สัมประสิทธิ์ และค่าผลต่างพลังงาน (ΔE)

7. การทดลองชุดที่เจ็ด (F7) 41 สัมประสิทธิ์ ประกอบด้วยค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมล (MFCC) จำนวน 13 สัมประสิทธิ์ ผลต่างค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมลลำดับที่หนึ่ง (ΔMFCC) จำนวน13 สัมประสิทธิ์ ผลต่างค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมลลำดับที่สอง (ΔΔMFCC) จำนวน13 สัมประสิทธิ์ ค่าผลต่างพลังงาน (ΔE)และค่าผลต่างพลังงานลำดับที่สอง (ΔΔE)

8. การทดลองชุดที่แปด (F8) 42 สัมประสิทธิ์ ประกอบด้วยค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมล (MFCC) จำนวน 13 สัมประสิทธิ์ ผลต่างค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมลลำดับที่หนึ่ง (ΔMFCC) จำนวน13 สัมประสิทธิ์ ผลต่างค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมลลำดับที่สอง (ΔΔMFCC) จำนวน13 สัมประสิทธิ์ ค่าพลังงงาน (E)ค่าผลต่างพลังงาน (ΔE)และค่าผลต่างพลังงานลำดับที่สอง (ΔΔE)

ตารางที่ 5 แสดงรายละเอียดค่าคุณลักษณะสำคัญ

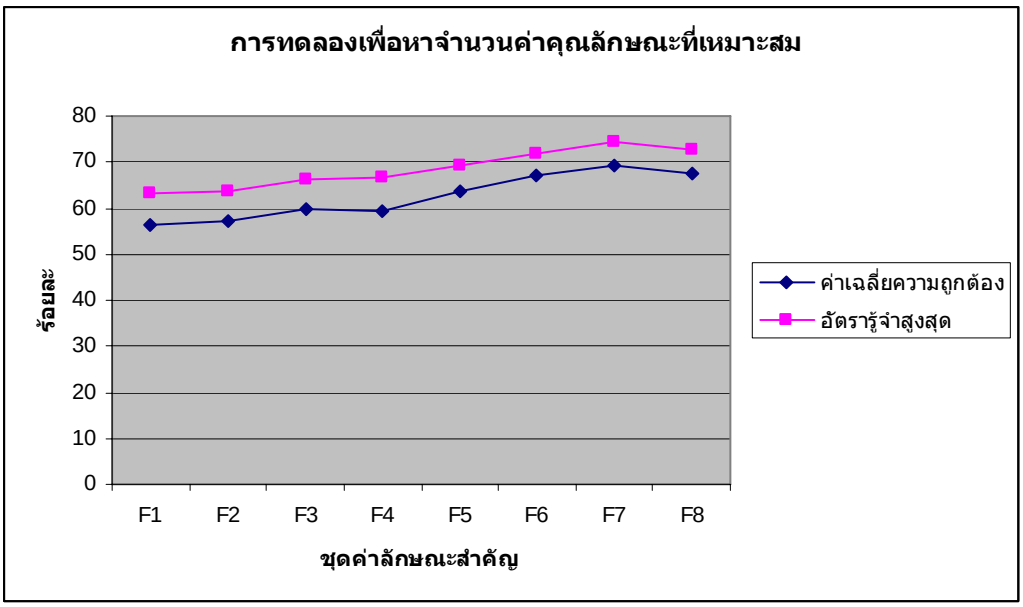
ค่าคุณลักษณะสำคัญที่	ชนิดของค่าคุณลักษณะสำคัญ	จำนวน
1	ค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมล (MFCC)	13
2	ผลต่างค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมลลำดับที่หนึ่ง (ΔMFCC)	13
3	ผลต่างค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมลลำดับที่สอง (ΔΔMFCC)	13
4	ค่าพลังงาน (E)	1
5	ค่าผลต่างพลังงาน (ΔE)	1
6	ค่าผลต่างพลังงานในลำดับที่สอง (ΔΔE)	1

ในแต่ละการทดลองนี้โดยในส่วนของแบบจำลองฮิดเดนมาร์คอฟในการรู้จำรูปแบบโดยแต่ละการทดลองนั้นเลือกใช้จำนวนสถานะที่ 5 และ 9 สถานะ รวมถึงมีการปรับเปลี่ยนค่ากลุ่มของเกาส์เซียน 4 ค่าคือ 2, 4, 6 และ 8 รวมเป็น 8 การทดลองย่อย ดังตัวอย่างตารางที่ 6

หลังจากนั้นทำการหาค่าเฉลี่ยความถูกต้องของแต่ละชุดการทดลอง เพื่อหาที่สูงสุด ผลลัพธ์ที่ได้คือการทดลองที่ 7 โดยให้ค่าอัตราความถูกต้องสูงสุดรวมถึงอัตราเฉลี่ยความถูกต้องสูงสุดดังแสดงในภาพที่ 15

ตารางที่ 6 ค่าความถูกต้องเมื่อใช้ค่าลักษณะสำคัญของเสียงที่ รวม 41 ค่า

	5 สถานะ	9 สถานะ
2 กลุ่มเกาส์เซียน	52.15	62.66
4 กลุ่มเกาส์เซียน	54.73	64.50
6 กลุ่มเกาส์เซียน	55.00	66.25
8 กลุ่มเกาส์เซียน	68.0	74.25



ภาพที่ 15 กราฟผลการรู้จำของแต่ละชุดค่าคุณลักษณะสำคัญ

จากผลลัพธ์ที่ได้ จำนวนของค่าลักษณะสำคัญที่ให้ผลลัพธ์ดีที่สุดคือ 5 คุณลักษณะ ประกอบด้วย ประกอบด้วยค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมล ผลต่างค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมลลำดับที่หนึ่ง ผลต่างค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมลลำดับที่สอง ค่าผลต่างพลังงาน และค่าผลต่างพลังงานลำดับที่สอง จำนวนรวม 41 ค่า

2. ผลการทดลองเพื่อหาจำนวนสถานะและจำนวนกลุ่มของเกาส์เซียนที่เหมาะสม

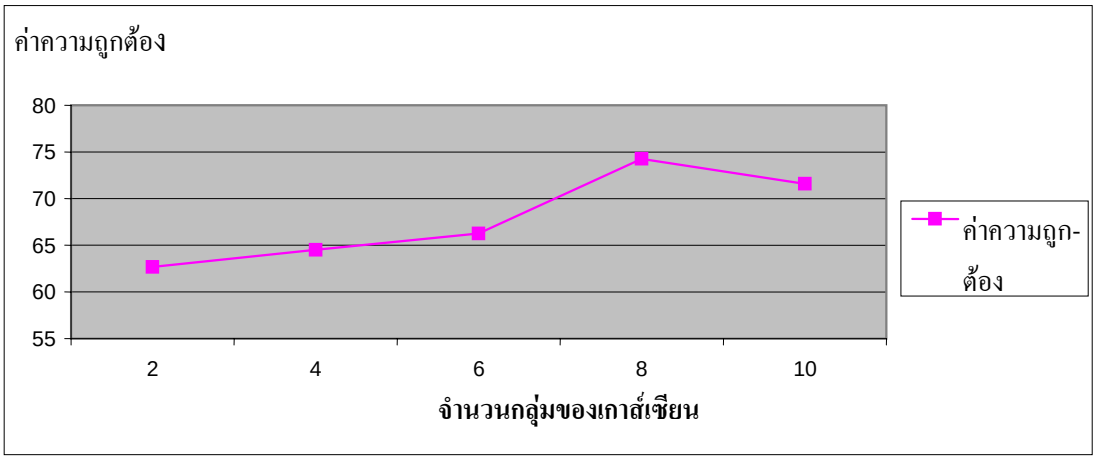
ในแบบจำลองฮิดเดนมาคอร์ฟแบบต่อเนื่องจำนวนสถานะของแบบจำลองและจำนวนกลุ่มเกาส์เซียนมีผลต่อความถูกต้องรู้จำของระบบ ในการทดลองนี้จึงหาค่าที่สถานะของ

แบบจำลองและจำนวนกลุ่มเกาส์เซียนที่ให้ผลลัพธ์ที่ดีที่สุด จากผลการทดลองที่ 1 ที่ผ่านมา สามารถสรุปค่าลักษณะสำคัญที่ให้ผลดีที่สุดมาคือ 5 คุณลักษณะจำนวน 41 ค่า ในการทดลองนี้จึงเลือกใช้ค่าคุณลักษณะจำนวน 41 ค่าสำหรับส่วนการดึงค่าคุณลักษณะสำคัญของเสียง ในส่วนการหาค่ากลุ่มของเกาส์เซียนที่เหมาะสมนั้น เริ่มจาก จำนวน 2 กลุ่มเกาส์เซียน และเพิ่มค่าทีละสอง ผลที่ให้ค่าความถูกต้องสูงสุดคือจำนวนกลุ่มของเกาส์เซียนจำนวน 8 กลุ่มเกาส์เซียน ดังตารางที่ 7 และ กราฟในภาพที่ 16

นอกจากนี้ยังทำการปรับค่าจำนวนสถานะเพื่อให้ได้ค่าจำนวนสถานะที่เหมะสม โดยเริ่มจากสถานะที่ 5 ไป สถานะที่ 12 ค่าความถูกต้องเพิ่มขึ้นจากร้อยละ 68.05 เป็น ร้อยละ 74.25

ตารางที่ 7 อัตราการรู้จำเสียงต่อจำนวนกลุ่มของเกาส์เซียน

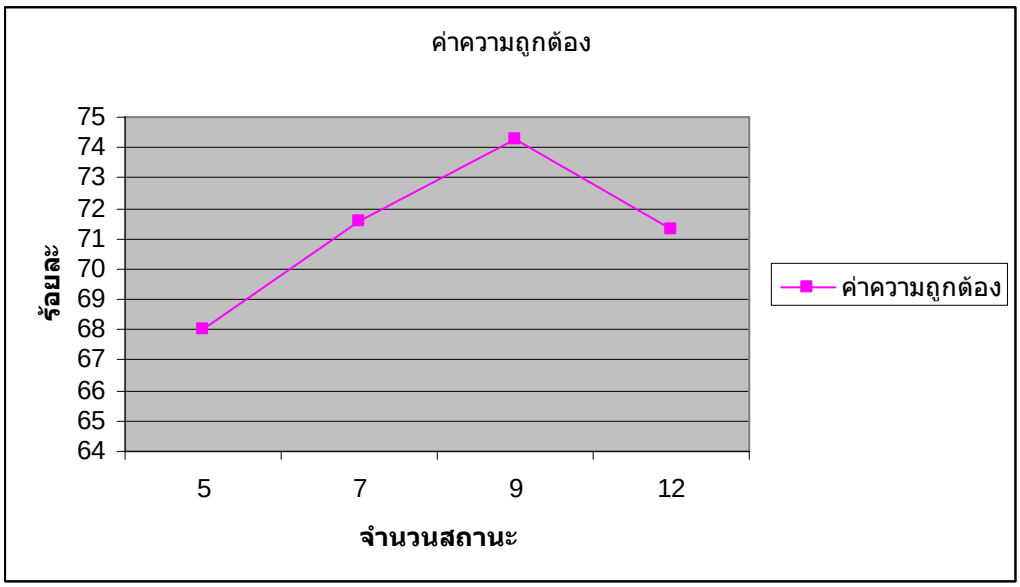
จำนวนกลุ่มของเกาส์เซียนที่ใช้	ค่าความถูกต้อง
2	62.66
4	64.50
6	66.25
8	74.25
10	71.60



ภาพที่ 16 กราฟผลการรู้จำของจำนวนกลุ่มเกาส์เซียน

ตารางที่ 8 อัตราการรู้จำเสียงต่อจำนวนสถานะ

จำนวนสถานะ	ค่าความถูกต้อง
5	68.05
7	71.60
9	74.25
12	71.33



ภาพที่17 กราฟผลการรู้จำของจำนวนสถานะที่เหมาะสม

หลังจากได้ค่าจำนวนสถานะของแบบจำลองและค่ากลุ่มเกาส์เซียนที่ให้ค่าสูงสุดแล้วในที่นี้คือ 9 สถานะต่อแบบจำลอง และ 8 กลุ่มเกาส์เซียนนำมาทดลองหาค่าพารามิเตอร์ที่เหมาะสมของค่าคุณลักษณะสำคัญของเสียงอีกครั้งได้ผลดังตารางที่ 9 ผลลัพธ์ที่ได้สูงสุดเมื่อใช้ 41 สัมประสิทธิ์ ประกอบด้วยค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมล (MFCC) จำนวน 13 สัมประสิทธิ์ ผลต่างค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมลลำดับที่หนึ่ง (ΔMFCC) จำนวน13 สัมประสิทธิ์ ผลต่างค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมลลำดับที่สอง (ΔΔMFCC) จำนวน13 สัมประสิทธิ์ ค่าผลต่างพลังงาน (ΔE)และค่าผลต่างพลังงานลำดับที่สอง (ΔΔE)

ตารางที่ 9 อัตราการรู้จำเสียงต่อจำนวนค่าคุณลักษณะสำคัญ

ลักษณะสำคัญที่ใช้	จำนวนของ ลักษณะสำคัญ	ผลการรู้จำ(%)
MFCC+ΔMFCC	26	63.33
MFCC+ΔMFCC+ΔE	27	63.54
MFCC+ΔMFCC+ΔE+ΔΔE	28	66.10
MFCC+ΔMFCC+E+ΔE+ΔΔE	29	66.60
MFCC+ΔMFCC+ΔΔMFCC	39	70.40
MFCC+ΔMFCC+ ΔΔMFCC+ΔE	40	71.80
MFCC+ΔMFCC+ΔΔMFCC+ ΔE+ΔΔE	41	**74.25**
MFCC+ΔMFCC+ΔΔMFCC+ E+ΔE+ΔΔE	42	72.66

4. ผลการทดลองจำนวนเสียงพูดที่ใช้ในการฝึกฝนระบบ

การทดลองเพื่อหาจำนวนเสียงที่เพียงพอสำหรับการฝึกฝนระบบในการรู้จำเสียงต่อเนื่องตัวเลขภาษาไทย ทำโดยวัดอัตราการรู้จำเสียงต่อเนื่องของผู้พูดทั้งหมด 100 คน แบ่งเป็นผู้ชาย 50 คนและ ผู้หญิง 50 คน (ดังตารางที่ 7) ในการทดลองแบ่งจำนวนเพศเท่าๆกันในการทดลองแต่ละครั้งโดยทดลองตั้งแต่ 20, 40, 60, 80 และ 100 เสียงผู้พูดตามลำดับ

ตารางที่ 10 อัตราการรู้จำเสียงต่อจำนวนเสียงที่ใช้ฝึกฝนระบบ

จำนวนเสียงผู้พูดที่ใช้ในการฝึกฝน	อัตรการรู้จำ (%)
20	58.18
40	62.40
60	69.54
80	72.56
100	74.25

จากผลการทดลองในตารางที่ 10 แสดงผลของจำนวนเสียงของผู้พูดที่ใช้ฝึกฝนกับ อัตราการรู้จำเสียงของระบบสูงขึ้นเมื่อเพิ่มผู้พูดที่ใช้ในการฝึกฝนจาก 20 ผู้พูดไปถึง 100 ผู้พูด โดยการเพิ่มจำนวนที่ 60 ผู้พูดขึ้นไปจะให้ผลความถูกต้องเพิ่มขึ้นอย่างชัดเจนที่สุดคือเพิ่มขึ้นร้อยละ 7.14 เมื่อเพิ่มจาก 40 ผู้พูดเป็น 60 ผู้พูด

5. <u>การวิเคราะห์ผลข้อผิดพลาดของระบบ</u>

การทดสอบประสิทธิภาพของระบบ ในส่วนนี้จะวิเคราะห์ค่าความผิดพลาดที่โดยทำการเพิ่มเสียงและแสดงให้เห็นที่สัดส่วนข้อผิดพลาดในแต่ละแบบ จากตารางที่ 11 จะเห็นว่าข้อผิดพลาดแบบแทรกและแบบแทนที่จะเกิดบ่อยมากที่สุด

ตารางที่ 12 แสดงการแทนที่ของตัวเลขที่เกิดขึ้นจากการทดลอง แสดงให้เห็นข้อผิดพลาดที่เกิดขึ้นในบางตัวเลขเช่น เลข 7 ถูกแทนที่ได้จำนวน 4 ตัวเลขคือเลข 1,6,8 และ 9 รูปแบบที่ผิดพลาดบ่อยมากที่สุดคือความผิดพลาดของกลุ่มเลข 3, 2 และ0 ถัดมาคือรูปแบบตัวเลข5 และ 9

<u>ตารางที่ 11</u> อัตราการข้อผิดพลาดชนิดต่างๆ

จำนวนผู้พูด	ข้อผิดพลาดแบบแทรก (%)	ข้อผิดพลาดแบบแทนที่ (%)	ข้อผิดพลาดแบบลดลง (%)
5	41.18	58.82	0
10	44.54	55.46	0
15	50.71	45.96	3.33
20	48.54	50.21	1.25
30	38.22	59.42	2.36
40	52.24	47.76	0

ตารางที่ 12 ผลความผิดพลาดในการแทนที่ของแต่ละตัวเลข

ตัวเลขที่ถูกต้อง	ตัวเลขที่ถูกแทนที่
0	2 , 3 , 10
1	2,7,8
2	0,3,6
3	0,2,5
4	0,3
5	9 , 3
6	1,2,5
7	1,6,8,9
8	1,2,3,9
9	3,5

สรุป

งานวิจัยนี้ได้เสนอระบบรู้จำเสียงพูดตัวเลขภาษาไทยต่อเนื่องแบบไม่ขึ้นกับบุคคล โดยทำการทดลองโดยการเปลี่ยนจำนวนองค์ประกอบของการดึงค่าลักษณะสำคัญของเสียง รวมถึงจำนวนค่าเกาส์ เซียน ตัวเลขที่ทดสอบความยาว 7 จากผลการทดลองจำนวน 40 ผู้พูด ผลการรู้จำที่ได้ดีพอสมควรในอัตราการรู้จำคำ ในการทดสอบหลังจากทำการเพิ่มค่าลักษณะสำคัญมีผลต่อค่าความถูกต้องของระบบอย่างชัดเจนนั่นคือค่าการเปลี่ยนแปลงของเซปสตรัมในอันดับที่ 1 และ อันดับที่ 2 รวมถึงค่าพลังงาน มีผลต่อประสิทธิภาพของระบบ การทำงานโดยใช้แบบจำลองฮิดเดนมาร์คอฟแบบต่อเนื่อง โดยจำนวนค่าคุณลักษณะสำคัญที่ใช้คือ 41 ค่า จำนวน สถานะ9 สถานะ โดยใช้จำนวนเกาส์เซียน 8 ค่า

อย่างไรก็ตามในการรู้จำเสียงต่อเนื่องของระบบ พบข้อผิดพลาดในกรณีที่เสียงที่นำมารู้จำเป็นคำที่มีเสียงคล้ายกันในเรื่องความยาวของเสียง หลายๆตัวเลขมีความยาวเสียงที่ใกล้เคียงกัน ในการแก้ไขควรจะนำข้อมูลทางด้านสัทศาสตร์เข้ามาช่วยในการรู้จำ นอกจากนี้ในการใช้เครื่องมือด้านการรู้จำเสียงเช่น เอชทีเค (HTK) ของมหาวิทยาลัยเคมบริดจ์ (Young, 2000) จะช่วยให้สามารถทดลองได้รวดเร็วขึ้นและหลายรูปแบบการทดลองกว่าการสร้างระบบขึ้นเอง

เอกสารและอ้างอิง

ทวิ ประทุมทาน. 2530. **การตรวจรู้เสียงพูดภาษาไทยโดยใช้หน่วยพยางค์.** วิทยานิพนธ์ปริญญาโท. จุฬาลงกรณ์มหาวิทยาลัย, กรุงเทพฯ.

ธีระ ภัทราพรนันท์. 2538. **การรู้จำเสียงพูดสระภาษาโดดๆ ไม่ขึ้นกับผู้พูดโดยการวัดสเปกตรัมดิสแตนซ์และไดนามิกไทม์วาร์ปปิง**. วิทยานิพนธ์ปริญญาโท. จุฬาลงกรณ์มหาวิทยาลัย, กรุงเทพฯ.

ประดิษฐ์พงษ์. 2536. ระบบ**จดจำคำพูดแบบไม่ต่อเนื่อง และไม่ขึ้นกับผู้พูดโดยใช้เทคนิคการจัดกลุ่ม**. วิทยานิพนธ์ปริญญาโท. สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ, กรุงเทพฯ.

พิชัย ชูกาญจนพิทักษ์ ไชยันต์ สุวรรณชีวะศิริ และมนูญ พ่วงพูล. 2002. การรู้จำเสียงพูดภาษาไทยต่อเนื่องจำนวนคำศัพท์ปานกลางเฉพาะบุคคล, น. 402-407. ใน **รายงานการประชุมวิชาการของ The 6th National Computer Science and Engineering Conference.**

ฐานียา สัตยพาณิช . 2544. **ระบบรู้จำเสียงภาษาไทยต่อเนื่องแบบเฉพาะบุคคลสำหรับการใช้งานอีเมล์.** วิทยานิพนธ์ปริญญาโท. มหาวิทยาลัยเกษตรศาสตร์, กรุงเทพ

Cosi, P., J. Hosom, and F. Ravera. 2000. High performance Italian continuous digit recognition, pp. 242–245. *In* **Proceeding of International Conference on Spoken Language Processing**. Beijing, China.

Cox, R. V., C. A. Kamm, L. R. Rabiner, and J. G. Wilpon. 2000. Speech and language processing for next-millennium communications services. *In* **Proceedings of the IEEE**, 88(8):1314–1337.

Davis, S. and P. Mermelstein. 1980. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. **IEEE Trans. on Acoustics, Speech, and Signal Processing,** 28:357–366.

Furui, S. 1986. Speaker-independent isolated word recognition using dynamic features of speech spectrum**. IEEE Trans. on ASSP**., 34(1):52–59.

_____. 1989. **Digital Speech Processing, Synthesis, and Recognition**. Marcel Dekker, Inc., New York.

Haeb-Umbach, R., D. Geller, and H. Ney. 1993. Improvements in connected digit recognition using linear discriminant analysis and mixture densities, pp. 349–352. ***In*** **Proceedings of the IEEE Int. Conference on Acoustics, Speech, and Signal Processing**. USA.

Jacobsen, C. N. and J. G. Wilpon. 1993. Automatic recognition of Danish natural numbers for telephone applicaitions, pp. 459–462. *In* **Proceedings of the IEEE Int. Conference on Acoustics, Speech, and Signal Processing**. USA.

Kasemsiri, W., C. Kimpan, and R. Kongkachandra. 2000. Refined language modeling for Thai continuous speech recognition, pp. 252–261. *In* **Proceedings of the 4th Symposium on Natural Language Processing**, Chiangmai, Thailand.

Kawai, H. and N. Higuchi. 1998. High performance Italian continuous digit recognition, pp. 341–344. *In* **Proceeding of the 5th International Conference on Spoken Language Processing.** Sydney, Australia.

Lee, K.-F. 1989. **Automatic Speech Recognition: The Development of the SPHINX System.** Marcel Dekker, Inc., New York.

Normandin, Y., R. Cardin, and R. D. Mori. 1994. High-perfomance connected digit recognition using maximum mutual information estimation. **IEEE Trans. on Acoustics, Speech, and Signal Processing**, 2(2):299–311.

Pensiri, R. and S. Jittapankul. 1995. Speaker-independent Thai numerical voice recognition by dynamic time warping, pp. 977–981. *In* **Proceedings of the 18th Electrical Engineering Conference**.

Pornsukchandra, W. 1996. **Speaker-Independent Thai numeral speech recognition using LPC and the back propagation neural network**. Master's thesis, Chulalongkorn Univ., Bangkok, Thailand, Department of Electrical Engineering. (In Thai).

Potisuk, S., M. Harper, and J. Gandour. May, 1995. Speaker-independent automatic classification of Thai tones in connected speech by analysis-synthesis method, pp. 632–635. *In* **Proceeding of International Conference on Acoustics, Speech, and Signal Processing**. MI, USA.

Rabiner, L. R. 1989. A tutorial on hidden markov models and selected applications in speech recognition. **Proceedings of the IEEE**, 77(2):257–286.

Rabiner, L. R., J. G. Wilpon, and F. K. Soong. 1989. High performance connected digit recognition using hidden markov models**. IEEE Trans. On Acoustics, Speech, and Signal Processing**, 37:1214–1225.

Rabinber, L. R. and B.-H. Juang. 1993. **Fundamentals of Speech Recognition. Prentice-Hall**, New Jersey.

Ramesh, P., J. G. Wilpon, M. A. McGee, D. B. Roe, C.-H. Lee, and L. R. Rabiner. 1992. Speaker-independent recognition of spontaneously spoken connected digits. **Speech Communication**, 11(2–3):229–235.

Shyu, R.-C., J.-F. Wang, and C.-C. Huang. 1993. Combining multi-section baysian template with level-building algorithm for robust connected

Mandarin digit recognition, pp. 213–217. ***In*** **Proceedings of International Symposium on VLSI Technology, System, and Application**. Taipei, Taiwan.

Sricharoenchai, A.. 2002. Thai Vowel Phoneme Recognition based on Auditory Model. *In* **Proceeding of The 4th Symposium on Natural Language Processing(SNLP' 02).**

Thubthong, N. and B. Kijsirikul. 1999. Syllable-based connected Thai digit speech recognition using neural network and duration modeling, pp. 785–788. *In* **Proceeding of IEEE International Symposium on Intelligent Signal Processing and Communication Systems**. Phuket, Thailand.

_____. 2001. Tone recognition of continuous Thai speech under tonal assimilation and declination effects using half-tone model. **International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems**, 9(6):815–825.

Wilpon, J. G., C. Lee, and L. R. Rabiner. 1991. Improvements in connected digit recognition using higher order spectral and energy features, pp. 349–352. *In* **Proceedings of the IEEE Int. Conference on Acoustics, Speech, and Signal Processing.** San Francisco, CA, USA.

ภาคผนวก

ภาคผนวก ก

ตารางภาคผนวกที่ 1 สัญญลักษณ์แทนเสียงของตัวเลขแต่ละตัว

ตัวเลข	สัญญลักษณ์แทนเสียง
0	/s uu n/
1	/n v ng/
2	/s @@ ng/
3	/ s aa m/
4	/s ii/
5	/ h aa/
6	/h o k /
7	/c e t/
8	/p xx t/
9	/ k aw/

ตารางภาคผนวกที่ 2 ตัวเลขที่ใช้ในการฝึกฝนและสัญญลักษณ์ที่ใช้แทนเสียง

ตัวเลขที่ใช้ในการฝึกฝน	สัญญลักษณ์แทนเสียง
9989796	k aw : k aw : p xx t : k aw : c e t : k aw : h o k
1367789	n v ng : s aa m: h o k:c e t:c e t: p xx t: k aw
9584839	k aw : h aa : p xx t : s ii : p xx t : s aam : kaw
8077678	p xxt: s uu n: c et: c et: h ok: c et: p xxt
5066574	h aa: s uu n: h o k: h o k: h aa: c e t: s ii
1056473	n v ng: s uu n: h aa: h o k: s ii: c e t: s aa m
3046372	s aa m: s uu n: s ii: h o k: s aa m: c e t: s @@ ng
7170362	c e t: n v ng: c e t: s uu n: s aa m: h o k: s @@ ng
4026160	s ii: s uun: s @@ ng: h o k: n v ng: h o k: s uu n
8594930	p xx t: h aa: k aw: s ii: k aw: s aa m: s uu n
5352514	h aa: s aa m: h aa: s @@ ng: h aa: n v ng: s ii
4434241	s ii: s ii: s aa m: sii: s @@ ng: s ii: n v ng
3323138	s aa m: s aa m: s @@ ng: s aa m: n v ng: s aa m: p xx t
9212002	k aw: s @@ ng: n v ng: s @@ ng: s uu n: s uu n: s @@ ng
2768114	s @@ ng: c e t: h o k: p xx t: n v ng: n v ng: s ii
6082818	h o k: s uu n: p xx t: s @@ ng: p xx t: n v ng: p xx t
5015545	h aa: s uu n: n v ng: h aa: haa: sii: haa
8878697	p x x t: p xx t: c e t: p xx t: h o k: k aw: c e t
0123456	s uu n: n v ng: s @@ ng: s aa m: s ii : h aa: h o k
2291909	s @@ ng: s @@ ng: k aw: n v ng: k aw: s uu n: k aw

ตารางภาคผนวกที่ 3 ตัวเลขและสัญญลักษณ์ที่ใช้แทนเสียงที่ใช้ในการทดสอบ

ตัวเลขที่ใช้ในการฝึกฝน	สัญญลักษณ์แทนเสียง
9989796	k aw : k aw : p xx t : k aw : c e t : k aw : h o k
1367789	n v ng : s aa m: h o k:c e t:c e t: p xx t: k aw
9584839	k aw : h aa : p xx t : s ii : p xx t : s aam : kaw
8077678	p xxt: s uu n: c et: c et: h ok: c et: p xxt
5066574	h aa: s uu n: h o k: h o k: h aa: c e t: s ii
1056473	n v ng: s uu n: h aa: h o k: s ii: c e t: s aa m
3046372	s aa m: s uu n: s ii: h o k: s aa m: c e t: s @@ ng
7170362	c e t: n v ng: c e t: s uu n: s aa m: h o k: s @@ ng
4026160	s ii: s uun: s @@ ng: h o k: n v ng: h o k: s uu n
8594930	p xx t: h aa: k aw: s ii: k aw: s aa m: s uu n
5352514	h aa: s aa m: h aa: s @@ ng: h aa: n v ng: s ii
4434241	s ii: s ii: s aa m: sii: s @@ ng: s ii: n v ng
3323138	s aa m: s aa m: s @@ ng: s aa m: n v ng: s aa m: p xx t
9212002	k aw: s @@ ng: n v ng: s @@ ng: s uu n: s uu n: s @@ ng
2768114	s @@ ng: c e t: h o k: p xx t: n v ng: n v ng: s ii
6082818	h o k: s uu n: p xx t: s @@ ng: p xx t: n v ng: p xx t
5015545	h aa: s uu n: n v ng: h aa: haa: sii: haa
8878697	p x x t: p xx t: c e t: p xx t: h o k: k aw: c e t
0123456	s uu n: n v ng: s @@ ng: s aa m: s ii : h aa: h o k
2291909	s @@ ng: s @@ ng: k aw: n v ng: k aw: s uu n: k aw

ภาคผนวก ข

การหาจำนวนค่าคุณลักษณะที่เหมาะสมของระบบมีรายละเอียดการเลือกใช้องค์ประกอบต่างๆดังนี้

1. การทดลองชุดที่หนึ่ง (F1) 26 สัมประสิทธิ์ ประกอบด้วยค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมล (MFCC) จำนวน 13 สัมประสิทธิ์ และผลต่างค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมลลำดับที่หนึ่ง (ΔMFCC) จำนวน13 สัมประสิทธิ์

2. การทดลองชุดที่สอง (F2) 27 สัมประสิทธิ์ ประกอบด้วยค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมล (MFCC) จำนวน 13 สัมประสิทธิ์ ผลต่างค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมลลำดับที่หนึ่ง (ΔMFCC) จำนวน13 สัมประสิทธิ์ และค่าผลต่างพลังงาน (ΔE) 1 สัมประสิทธิ์

3. การทดลองชุดที่สาม (F3) 28 สัมประสิทธิ์ ประกอบด้วยค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมล (MFCC) จำนวน 13 สัมประสิทธิ์ และผลต่างค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมลลำดับที่หนึ่ง (ΔMFCC) จำนวน13 สัมประสิทธิ์ ค่าผลต่างพลังงาน (ΔE)และค่าผลต่างพลังงานลำดับที่สอง (ΔΔE) อย่างละ 1 สัมประสิทธิ์

4. การทดลองชุดที่สี่ (F4) 29 สัมประสิทธิ์ ประกอบด้วยค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมล (MFCC) จำนวน 13 สัมประสิทธิ์ และผลต่างค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมลลำดับที่หนึ่ง (ΔMFCC) จำนวน13 สัมประสิทธิ์ ค่าพลังงาน (E) ค่าผลต่างพลังงาน (ΔE) และค่าผลต่างพลังงานลำดับที่สอง (ΔΔE)

5. การทดลองชุดที่ห้า (F5) 39 สัมประสิทธิ์ ประกอบด้วยค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมล (MFCC) จำนวน 13 สัมประสิทธิ์ ผลต่างค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมลลำดับที่หนึ่ง (ΔMFCC) จำนวน13 สัมประสิทธิ์ และ ผลต่างค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมลลำดับที่สอง (ΔΔMFCC) จำนวน13 สัมประสิทธิ์

6. การทดลองชุดที่หก (F6) 40 สัมประสิทธิ์ ประกอบด้วยค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมล (MFCC) จำนวน 13 สัมประสิทธิ์ ผลต่างค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมลลำดับที่

หนึ่ง (ΔMFCC) จำนวน13 สัมประสิทธิ์ ผลต่างค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมลลำดับที่สอง (ΔΔMFCC) จำนวน13 สัมประสิทธิ์ และค่าผลต่างพลังงาน (ΔE)

7. การทดลองชุดที่เจ็ด (F7) 41 สัมประสิทธิ์ ประกอบด้วยค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมล (MFCC) จำนวน 13 สัมประสิทธิ์ ผลต่างค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมลลำดับที่หนึ่ง (ΔMFCC) จำนวน13 สัมประสิทธิ์ ผลต่างค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมลลำดับที่สอง (ΔΔMFCC) จำนวน13 สัมประสิทธิ์ ค่าผลต่างพลังงาน (ΔE)และค่าผลต่างพลังงานลำดับที่สอง (ΔΔE)

8. การทดลองชุดที่แปด (F8) 42 สัมประสิทธิ์ ประกอบด้วยค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมล (MFCC) จำนวน 13 สัมประสิทธิ์ ผลต่างค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมลลำดับที่หนึ่ง (ΔMFCC) จำนวน13 สัมประสิทธิ์ ผลต่างค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมลลำดับที่สอง (ΔΔMFCC) จำนวน13 สัมประสิทธิ์ ค่าพลังงงาน (E)ค่าผลต่างพลังงาน (ΔE)และค่าผลต่างพลังงานลำดับที่สอง (ΔΔE)

ตารางภาคผนวกที่ 3 ค่าความถูกต้องเมื่อใช้ค่าลักษณะสำคัญของเสียง 26 ค่า

	5 สถานะ	9 สถานะ
2 กลุ่มเกาส์เซียน	52.15	55.42
4 กลุ่มเกาส์เซียน	54.73	56.80
6 กลุ่มเกาส์เซียน	55.00	59.75
8 กลุ่มเกาส์เซียน	55.45	62.20

ตารางภาคผนวกที่ 4 ค่าความถูกต้องเมื่อใช้ค่าลักษณะสำคัญของเสียง 27 ค่า

	5 สถานะ	9 สถานะ
2 กลุ่มเกาส์เซียน	53.24	57.36
4 กลุ่มเกาส์เซียน	54.88	58.88
6 กลุ่มเกาส์เซียน	56.30	62.25
8 กลุ่มเกาส์เซียน	57.25	63.50

ตารางภาคผนวกที่ 5 ค่าความถูกต้องเมื่อใช้ค่าลักษณะสำคัญของเสียง 28 ค่า

	5 สถานะ	9 สถานะ
2 กลุ่มเกาส์เซียน	56.24	59.76
4 กลุ่มเกาส์เซียน	58.72	60.54
6 กลุ่มเกาส์เซียน	60.41	62.15
8 กลุ่มเกาส์เซียน	61.96	65.10

ตารางภาคผนวกที่ 6 ค่าความถูกต้องเมื่อใช้ค่าลักษณะสำคัญของเสียง 29 ค่า

	5 สถานะ	9 สถานะ
2 กลุ่มเกาส์เซียน	58.42	60.77
4 กลุ่มเกาส์เซียน	59.28	61.50
6 กลุ่มเกาส์เซียน	61.05	64.85
8 กลุ่มเกาส์เซียน	63.23	66.60

ตารางภาคผนวกที่ 7 ค่าความถูกต้องเมื่อใช้ค่าลักษณะสำคัญของเสียง 39 ค่า

	5 สถานะ	9 สถานะ
2 กลุ่มเกาส์เซียน	62.06	63.80
4 กลุ่มเกาส์เซียน	64.28	65.50
6 กลุ่มเกาส์เซียน	65.05	68.85
8 กลุ่มเกาส์เซียน	67.17	70.40

ตารางภาคผนวกที่ 8 ค่าความถูกต้องเมื่อใช้ค่าลักษณะสำคัญของเสียง 40 ค่า

	5 สถานะ	9 สถานะ
2 กลุ่มเกาส์เซียน	62.86	63.95
4 กลุ่มเกาส์เซียน	65.20	66.35
6 กลุ่มเกาส์เซียน	66.05	69.12
8 กลุ่มเกาส์เซียน	67.05	71.80

ตารางภาคผนวกที่ 9 ค่าความถูกต้องเมื่อใช้ค่าลักษณะสำคัญของเสียง 42 ค่า

	5 สถานะ	9 สถานะ
2 กลุ่มเกาส์เซียน	63.87	65.50
4 กลุ่มเกาส์เซียน	66.40	66.35
6 กลุ่มเกาส์เซียน	67.87	71.27
8 กลุ่มเกาส์เซียน	69.47	72.66